The BD2K Concept Network:
Open Sharing of Active-Learning and Tools Online

The University of California Los Angeles

PI: Christopher Lee

The proposed project will create online services, teaching materials sharing, and training for instructors and students to 1) expand and tailor Big Data To Knowedge (BD2K) learning for new audiences in bioinformatics, medical informatics and biomedical applications; 2) use active- learning to greatly increase conceptual understanding and real-world problem-solving ability; 3) directly measure learning effectiveness; and 4) boost the number of students that successfully complete BD2K courses. Tailoring the core concepts for BD2K success to teach diverse biomedical audiences is crucial both because these interdisciplinary concepts are a key barrier to entry, and because they are vital for real-world BD2K problem-solving ability. The UCLA/UCSD project team will: 1) provide an open, online repository where BD2K instructors worldwide can find, author, and share peer-reviewed active-learning exercises such as concept tests (already over 600), and immediately use them in class (with students answering with their smartphones or laptops); 2) catalyze the development, usage and validation of candidate BD2K concept inventories for rigorously measuring learning gains, via an accelerated approach of open- response concept testing and online data collection; 3) provide BD2K instructors a collaborative, peer-reviewed sharing and remixing platform for active-learning materials such as algorithm projects, hands-on data mining projects (via convenient "cloud projects"), exercises and problems, as well as "courselet" recording tools that automatically record video and audio on the instructor's laptop while they teach; 4) provide students anywhere free online courselets each about one key BD2K concept, consisting of brief videos tightly integrated with concept tests and all the active- learning exercises described above, and designed as an online persistent-learning community unified by concepts, in which students learn from the community's consolidated error models (common errors for a specific BD2K concept), effective remediations and counter-examples for each error model. Testing of this instructional approach for 3 years has doubled successful student completions of a BD2K methods course at UCLA, by reducing attrition, while simultaneously increasing conceptual understanding (mean exam scores). This approach will also be disseminated by: 1) pilot projects with BD2K instructors at UCLA and partner institutions, with detailed evaluation studies to identify critical success factors; 2) workshops (both online and onsite) for training instructors how to teach effectively with these tools in their BD2K courses; 3 online services and courselets. PUBLIC HEALTH RELEVANCE: Big Data to Knowledge (BD2K) education means bringing sophisticated data mining skills and thinking to researchers and clinicians throughout the biomedical enterprise, a most challenging interdisciplinary learning curve. This cannot succeed without the kinds of hands-on learning exercises that are hard to find in BD2K textbooks, but that students need, such as data-mining projects with real datasets and real computational powertools, concept tests and concept inventories that rigorously teach and measure conceptual understanding, and algorithm projects where students prove their understanding of a challenge problem, by writing code that can correctly solve any test case thrown at it. We will provide BD2K instructors a collaborative, peer- reviewed sharing platform for immediately using all of these kinds of active-learning materials in class (currently containing over 2000 BD2K exercises and related materials), and BD2K students free online courselets each about one key BD2K concept, consisting of brief videos tightly integrated with concept tests and all the active-learning exercises described above.